Data Storage

Representing Text

Representing Text

Representation as Code

- ✓ Each Textual Symbol is represented with a unique bit pattern
- ✓ Normally 8 bits for each character
- ✓ "Virtual University"
- ✓ 18 characters = 18*8 (144) bits or 18 byte

Representing Text

Codes

- ✓ In 1940's and 1950's many such codes were designed
- ✓ American National Standards Institute (ANSI)
- ✓ American Standard Code for Information Interchange (ASCII

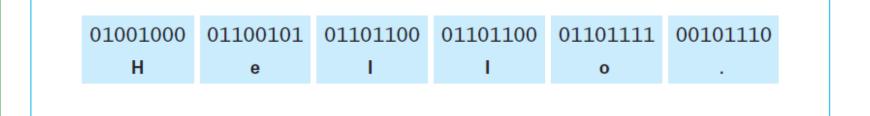
ASCII

- ✓ 7 bit for information and most significant bit has zero
- ✓ 27combinations= 128
- ✓ Uppercase, lower case, punctuation marks, digits 0 to 9, line feed, carriage returns, and tabs

ASCII codes

✓ Go to 577 of your book

ASCII codes



Limitations of ASCII codes

- ✓ Only 128 characters
- ✓ International Organization for Standardization (ISO) come-up with many extensions to ASCII
- ✓ One to support western language symbols.

Limitations of ASCII-extensions

Two Issues

- ✓ 256 are still insufficient to denote all language symbols
- ✓ Document having multiple languages could not be read as it should follow a one standard

Unicode

- ✓ Internationalization of codes by manufacturers of hardware and software
- ✓ Unique patterns of 21 bits
- ✓ Compliance with ASCII
- ✓ Supporting thousands of character sets of Chinese, Herbew, Japaneze,

UTF-8

- ✓ Uses 24 to 32 bits having large possibilities for expansion
- ✓ 224= 16,777,216 unique symbols
- ✓ File consisting of long symbols encoded with ASCII or Unicode is called text file.

Summary

Representing Text

- ✓ How text is represented in binary
- ✓ ASCII and Unicode